Global Z-score Normalization Enables Training-free Automatic Detection of Punctate White Matter Lesions in Neonates

Hoi Ching Wat[1,2], Erjun Zhang[1,3], Kylie Xu[1,4], Gregory Lodygensky[1,5,6]
[1]CHU Sainte-Justine Hospital Research Center, Montreal, QC, Canada
[2]Department of Biomedical Engineering, McGill University, Montreal, QC, Canada
[3]Neuropoly Lab, Institute of Biomedical Engineering, Polytechnique Montréal, Montreal, QC, Canada
[4]Department of Medicine, University of Montréal, Montreal, QC, Canada
[5]Department of Pediatrics, University of Montréal, Montreal, QC, Canada
[6]Department of Pharmacology and Physiology, University of Montréal, Montreal, QC, Canada

Keywords: Punctate White Matter Lesions, Neonates, Auto-Segmentation, Magnetic Resonance Imaging

Corresponding author email: hoi.wat@mail.mcgill.ca
Published: 03/25/2025
https://doi.org/10.26443/msurj.v21i2.506

Punctate white matter lesions (PWMLs) are common in preterm neonates and linked to adverse neurodevelopmental outcomes. Conventional identification relies on subjective visual contrast on MRI. We propose a global, training-free z-score normalization framework for automatic PWML detection, enabling standardized, interpretable segmentation across modalities without machine learning or subject-specific calibration. We analyzed 44 neonates from the dHCP dataset, comprising 1069 axial slices from T1w, T2w, and mean diffusivity (MD) maps. Reference masks were generated by consensus among four trained raters. Slices within each modality were standardized to a global z-score space, allowing comparable intensity distributions. For each modality, a detection threshold of z-score was swept from -3 to 3 standard deviations to construct receiver operating characteristic (ROC) curves. Lesion–white matter contrast ratios (CR) were also computed. Inter-rater reliability (Cohen's k) reached 0.88 for T1w, 0.77 for T2w, and 0.73 for MD. PWML–white matter contrast differences were statistically significant ($p < 0.01$). Contrast magnitude was greatest for T1w (CR = 0.57 ± 0.19), compared to T2w (-0.18 ± 0.08) and MD (-0.14 ± 0.12). ROC analysis revealed that T1w outperformed the other modalities, achieving an AUC of 0.98 at an optimal threshold of 0.84 (True Positive Rate = 0.90, False Positive Rate = 0.06). In contrast, T2w and MD achieved lower AUCs of 0.84 and 0.67, respectively. Global z-score normalization yields a robust, fully automatic lesion detection method. This interpretable baseline achieves high sensitivity with minimal false positives and provides a reproducible benchmark for evaluating deep-learning or quantitative MRI pipelines.